Coursera Capstone Project

Predicting the Severity of a Potential Accident

Introduction

- This data set is about accident (car collisions) severity. This data includes all types of collisions.
- Collisions will display at the intersection or mid-block of a segment.
- The data dates weekly from 2004 to present.
- The data has been collected from the Seattle Department of Transportation.

Problem

- Data can be used to conclude what influences car accidents.

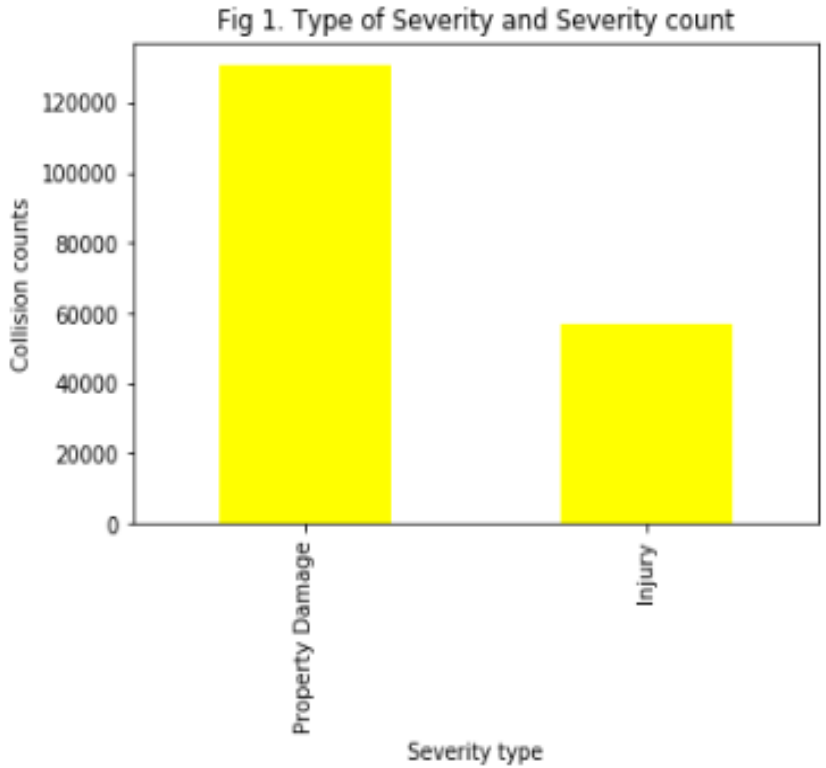
Data acquisition and cleaning

- Most important part of the data analysis is data cleaning and understanding.
- Severity of the accidents was segmented mainly into 2 groups:

    1) collisions which only involved property damage;

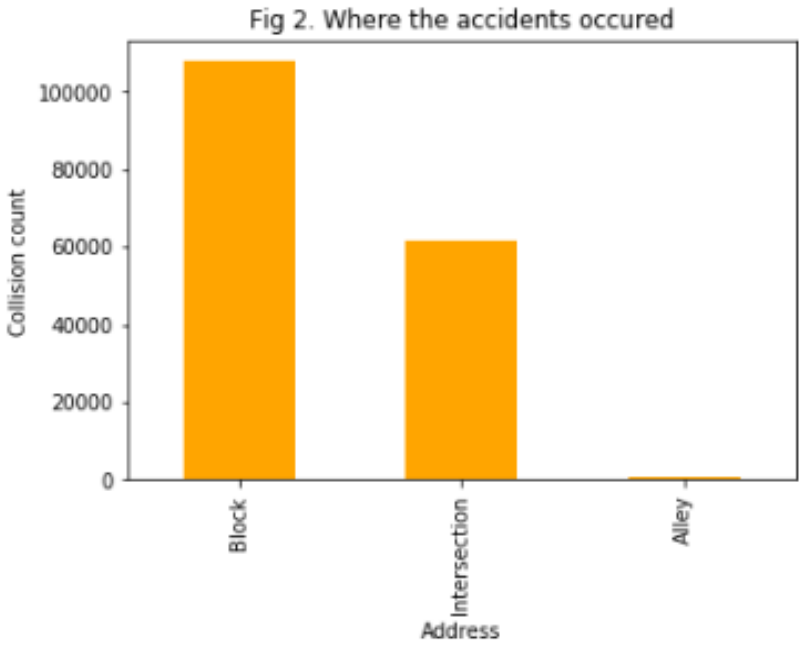
    2) collisions which involved injury.

Relationship between severity types



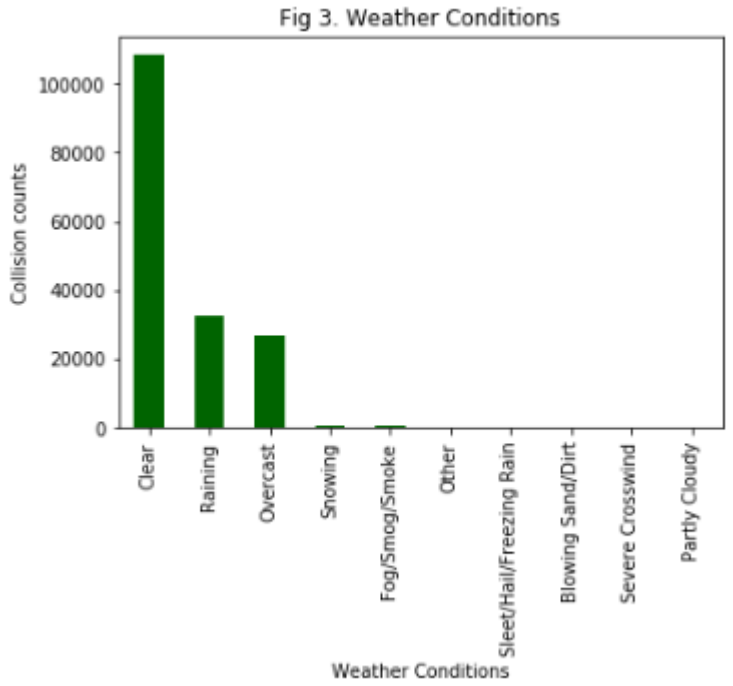
From the Fig.1 we can clearly see that in most of collision cases property was damaged. Property damage was in 136485 collisions, injuries was in 58188 collisions. That clearly illustrates Fig. 1.

Relationship between address and collision count



From the Fig.2 we can see that most of the accidents occured in the blocks, less accident occured at the intersections. Least accidents occured at the alley.

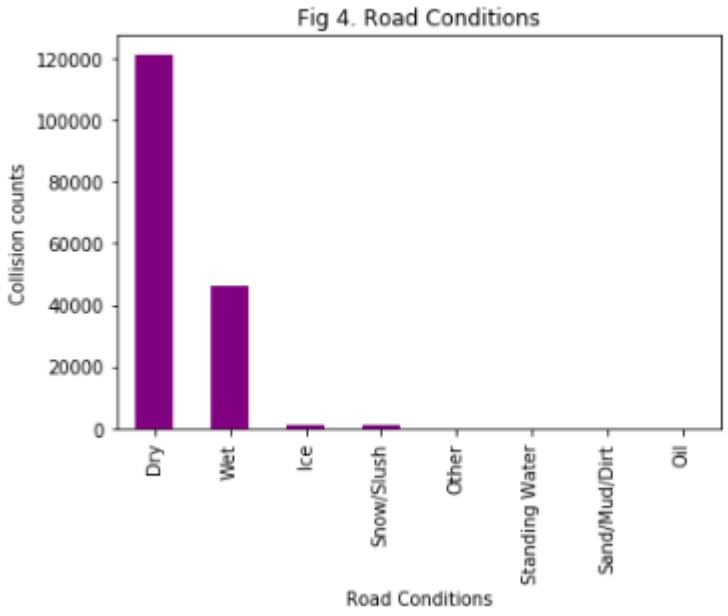
Relationship between weather conditions and collision count



Clear	108507
Raining	32599
Overcast	26863
Snowing	827
Fog/Smog/Smoke	549
Other	253
Sleet/Hail/Freezing Rain	110
Blowing Sand/Dirt	43
Severe Crosswind	25
Partly Cloudy	5

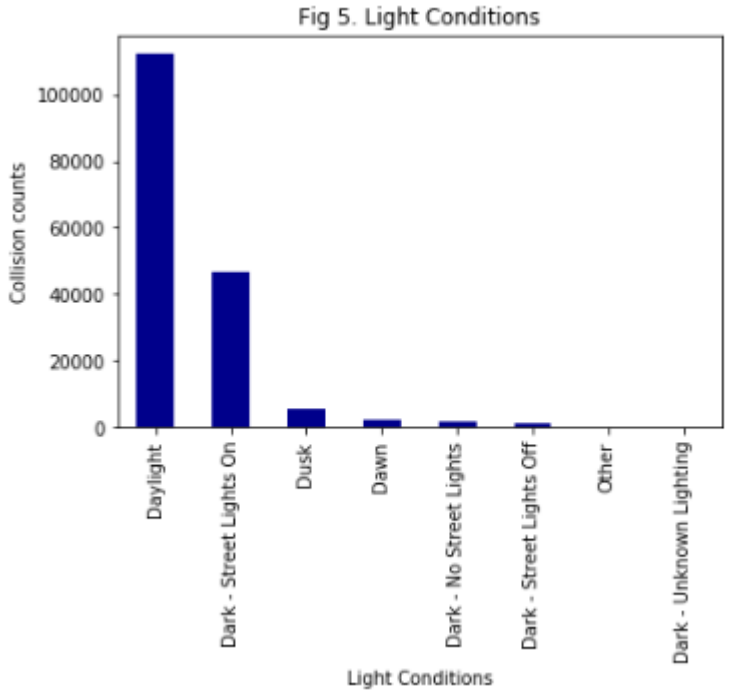
From the Fig.3 we can see that most of the accidents occured in the good weather conditions. (Most accidents happened in the 'clear' weather condition.)

Road condition influence to accidents



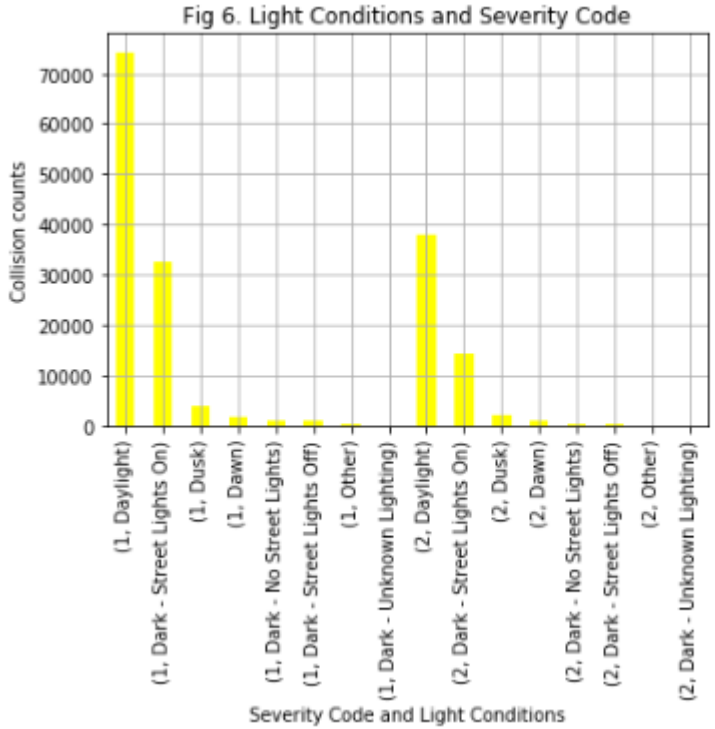
In Fig. 4 we can see that most of collisions happened at the dry road conditions.

Light Conditions influence to accidents



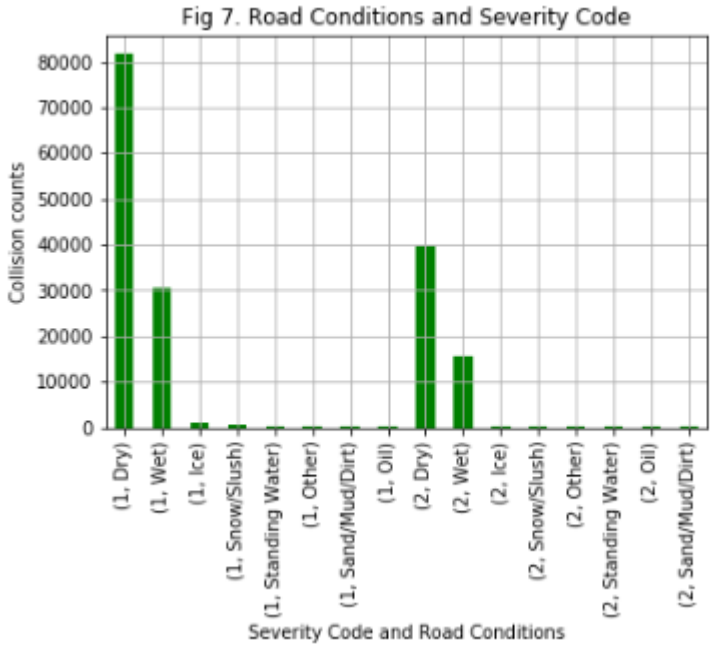
From the Fig. 5. we can see that most accidents happened in the daylight. That can happen because of most of the traffic happens also in the daylight. People goes to jobs, schools and so on.

Light Conditions influence to accidents



- Majority of the accidents took place in daylight (property damage and injuries) and in dark (with street lights on).
- This may conclude that most of accidents happens not because of daytime, but because of that people are unaware of situation, not paying full attention.

Road Conditions influence to accidents



Most of car accidents where property was damaged or people got injuried, took place in dry road condition.

Predictive Modeling

- There are two types of models, regression and classification, that can be used to predict player improvement. Regression models can provide additional information on the amount of improvement, while classification models focus on the probabilities a player might improve.
- The underlying algorithms are similar between regression and classification models, but different audience might prefer one over the other.
- I applied Linear Regression, Support Vector Machine, K –Neariest Number, Decision Tree models.

	Linear Regression	Support Vector Machine	K –Neariest Number	Decision Tree
Train set Accuracy	0.6706	0.6706	0.6439	0.6706
Test set Accuracy	0.6749	0.6749	0.6486	0.6749

Conclusion

- Most of accidents happened in the daylight, dry road conditions. That concludes that most of accidents happens because of human influence (not paying attention, lack of sleep and so on).
- Data was cleaned and prepared for data analysis and model building.
- The four models we built are all very similar in terms of prediction and accuracy.
- The highest prediction accuracy is about 67.49%.
- Most accurate model was "Suppor Vector Machine", "Logistic Regression" and "Decision Tree".
- In this project, we have found the major environmental factors and road conditions that affect car accidents. Also we found a building a model that can help predict the severity of car accidents based on these conditions.
- Based on the data analysis and results, we can make some recommendations to improve the safety of drivers, pedestrians and others. Most helpful advice would be to pay attention and to watch the signs.